# Matrix Element Method in HEP: Transfer Functions, Efficiencies, and Likelihood Normalization

Igor Volobouev

*i.volobouev@ttu.edu*

Texas Tech University, Department of Physics, Box 41051,
Lubbock, Texas, USA 79409

January 13, 2011

**Abstract**

This article surveys the procedures used for deriving detector transfer functions and normalizing probability densities for the statistical analysis technique known as the "matrix element method" in the context of high energy physics (HEP) data analysis. Common misconceptions about transfer functions and efficiencies are pointed out and clarified.

## 1 Matrix Element Analysis in a Nutshell

The matrix element (ME) method was introduced into the HEP data analysis practice relatively recently, in a pioneering measurement of the top quark mass by the D0 Collaboration [1, 2]. The method can be utilized both in searches of new phenomena and in precision measurements of physical model parameters. It consists in calculating the probability of observing an event in a particle detector according to

$$P_{\mathrm{ev}}(\mathbf{y}|\mathbf{a}) = \sum_i f_i P_i(\mathbf{y}|\mathbf{a}) \tag{1}$$

where $\mathbf{y}$ is the vector of observed quantities in the measurement space $Y$, $\mathbf{a}$ is the vector of model parameters (both theoretical and instrumental), and $f_i$ are fractions of different non-interfering production channels consistent with $\mathbf{y}$. In the following discussion, it will be assumed that $f_i \geq 0$ for every production channel $i$ and that the constraint $\sum_i f_i = 1$ is imposed. For each channel $i$, the probability to measure $\mathbf{y}$ is estimated from

$$P_i(\mathbf{y}|\mathbf{a}) = \frac{\Omega(\mathbf{y})}{\sigma_i(\mathbf{a}) A_i(\mathbf{a})} \int_{X_i} W_i(\mathbf{y}|\mathbf{x},\mathbf{a})\, \epsilon_i(\mathbf{x},\mathbf{a})\, |M_i(\mathbf{x},\mathbf{a})|^2\, T_i(\mathbf{x},\mathbf{a})\, d\mathbf{x} \tag{2}$$

where

$\Omega(\mathbf{y})$ — Indicator function for the analysis acceptance (1 for events which pass the event selection criteria, 0 otherwise). This term can be replaced by 1 in case only accepted events are considered.

$\mathbf{x}$ — Variables which uniquely specify a point in the channel phase space $X_i$, $\mathbf{x} \in X_i$.

$d\mathbf{x}$ — Differential element of the phase space $X_i$.

$\sigma_i(\mathbf{a})$ — Channel cross section: $\sigma_i(\mathbf{a}) = \int_{X_i} |M_i(\mathbf{x},\mathbf{a})|^2\, T_i(\mathbf{x},\mathbf{a})\, d\mathbf{x}$.

$A_i(\mathbf{a})$ — Overall experimental acceptance.

$W_i(\mathbf{y}|\mathbf{x},\mathbf{a})$ — Detector transfer function. This is the probability density to observe detector response $\mathbf{y} \in Y$ when the "true" phase space coordinate of the event is $\mathbf{x}$. This function is normalized by $\int_Y \Omega(\mathbf{y}) W_i(\mathbf{y}|\mathbf{x},\mathbf{a})\, d\mathbf{y} = 1$ for every value of $i$, $\mathbf{x}$, and $\mathbf{a}$.

$\epsilon_i(\mathbf{x},\mathbf{a})$ — Efficiency to detect an event originated at the phase space point $\mathbf{x}$ (*i.e.,* the probability that an event originated at $\mathbf{x}$ actually ends up with $\mathbf{y}$ for which $\Omega(\mathbf{y}) = 1$).

$|M_i(\mathbf{x},\mathbf{a})|^2$ — Squared matrix element of the process.

$T_i(\mathbf{x},\mathbf{a})$ — Other factors which do not depend on $\mathbf{y}$ (*e.g.,* flux of colliding beams, parton distribution functions).

The normalization condition imposed on $W_i(\mathbf{y}|\mathbf{x},\mathbf{a})$ ensures that $P_i(\mathbf{y}|\mathbf{a})$ is a properly normalized probability density in $Y$ for all values of $\mathbf{a}$. Indeed,

$$\int_Y P_i(\mathbf{y}|\mathbf{a})\, d\mathbf{y} = \int_Y \frac{\Omega(\mathbf{y})}{\sigma_i(\mathbf{a}) A_i(\mathbf{a})} \int_{X_i} W_i(\mathbf{y}|\mathbf{x},\mathbf{a})\, \epsilon_i(\mathbf{x},\mathbf{a})\, |M_i(\mathbf{x},\mathbf{a})|^2\, T_i(\mathbf{x},\mathbf{a})\, d\mathbf{x}\, d\mathbf{y}.$$

Integrating over $\mathbf{y}$ first and taking into account the normalization of the transfer function, one gets

$$\int_Y P_i(\mathbf{y}|\mathbf{a})\, d\mathbf{y} = \frac{1}{\sigma_i(\mathbf{a}) A_i(\mathbf{a})} \int_{X_i} \epsilon_i(\mathbf{x},\mathbf{a})\, |M_i(\mathbf{x},\mathbf{a})|^2\, T_i(\mathbf{x},\mathbf{a})\, d\mathbf{x} = \frac{\langle \epsilon_i(\mathbf{x},\mathbf{a}) \rangle}{A_i(\mathbf{a})} \equiv 1,$$

where the symbol $\langle ... \rangle$ stands for averaging with respect to the process density $p(\mathbf{x}|\mathbf{a})$ in the phase space: $p(\mathbf{x}|\mathbf{a}) = \frac{1}{\sigma_i(\mathbf{a})} |M_i(\mathbf{x},\mathbf{a})|^2\, T_i(\mathbf{x},\mathbf{a})$. With properly normalized $P_i(\mathbf{y}|\mathbf{a})$, normalization of $P_{\rm ev}(\mathbf{y}|\mathbf{a})$ is ensured as well.

The formulation of the ME method just described assumes that the production channel fractions $f_i$ are known with sufficient precision and are not of interest in the data analysis. It can be easily generalized for use with some densities $\pi(f_i|\mathbf{a})$ which represent the prior knowledge about $f_i$ as well as to the case in which $f_i$ are parameters to be measured.

The ME approach offers several important advantages over all other data analysis schemes commonly used in HEP:

— The method is universal and can be applied to a wide variety of particle processes for which theoretical models have been established.

— The theoretical assumptions about the process under study (matrix elements $M_i(\mathbf{x},\mathbf{a})$, channel fractions $f_i$, parton distribution functions) are incorporated into the data analysis in the most efficient manner. The whole procedure can be viewed as a Bayesian marginalization of the event probability over all unobserved degrees of freedom. Particle theory provides a well-motivated informative prior for this marginalization.

— Some widely used data analysis methods introduce implicit assumptions about the shape of detector resolution functions. In particular, methods based on $\chi^2$ minimization (used, *e.g.,* in kinematic fitting) assume Gaussian measurement errors, and the effect of this assumption on the quality of statistical modeling can not be quantified within the $\chi^2$-based method itself. There is no such inherent restriction in the ME approach; very detailed and precise detector models can be usefully employed.

— Maximization of the likelihood $L(\mathbf{a}) = \prod P_{\rm ev}(\mathbf{y}|\mathbf{a})$ results in an efficient (in the statistical sense) estimate of the parameter $\mathbf{a}$. Straightforward profiling or marginalization of the likelihood can be utilized in case some of the $\mathbf{a}$ dimensions are not of interest and can be treated

as nuisance parameters. In effect, systematic uncertainties (which are nothing else but the uncertainties due to imprecisely known values of nuisance parameters) are calculated *on the event-by-event basis* and, therefore, each event contributes into the overall parameter estimate with an optimal weight which takes into account both statistical and systematic uncertainty.

— According to the Neyman-Pearson lemma, the likelihood ratio

$$r_j(\mathbf{y}|\mathbf{a}) = \frac{P_j(\mathbf{y}|\mathbf{a})}{\sum_{i \neq j} f_i P_i(\mathbf{y}|\mathbf{a})}$$

is the optimal discriminant function for channel $j$. In particular, when channel $j$ represents the signal under study, this function can be used to assign events to either "signal" or "background" category by testing whether $r_j(\mathbf{y}|\mathbf{a}) >$ cutoff. This method achieves the best misclassification rate among all possible classifiers with the same signal selection efficiency.

Despite all these advantages, due to its complexity and lack of standard computational tools the ME approach is not necessarily the obvious first choice among various HEP data analysis techniques. Consider, for example, the most general case in which the integration region $X_i$ in Eq. 2 corresponds to the relativistic phase space of all initial and final state particles, and the space of measurements $Y$ corresponds to the complete data record produced by an experimental apparatus in a triggered high energy event. This general case will remain intractable for any foreseeable future because the dimensionality of the integral and the complexity of the integrand are just too high: thousands of particles can be produced in a high energy collision and then traced in a detector with millions of data acquisition channels. Reducing the dimensionality of $X_i$ and $Y$ spaces to a manageable level while preserving the information about parameters of interest is one of the most critical issues in practical applications of the ME method.

## 2 A Comment on Dimensionality Reduction

The first and the most important dimensionality reduction stage employed by all collider experiments is the process of "event reconstruction", and "jet reconstruction" in particular. In all HEP applications of the ME method developed so far, the final state "soft QCD" processes (parton showering and hadronization) were combined together with the detector response, and this combination was subsequently modeled empirically by the jet transfer functions. This particular approach is well motivated (at least at high c.m.s. energies such as those of the Tevatron and the LHC) by the QCD factorization theorem and results in a drastic dimensionality reduction of $X_i$ which then becomes the phase space of parton-level quantities. In addition to jets, the event reconstruction procedure also produces other high-level "physics objects" (lepton candidates, decay vertices, missing transverse energy, *etc.*), so the measurement space $Y$ is typically associated with the variables which describe such objects.

After this first and necessary dimensionality reduction stage, calculation of the Eq. 2 integral can still remain a formidable problem. For example, the leading order parton phase space of the reaction $p\bar{p} \to t\bar{t} \to W^+bW^-\bar{b} \to \ell\nu + 4$ jets, $\ell = e$ or $\mu$ (this is the "golden channel" for measuring the top quark mass at the Tevatron) is 24-dimensional[1], while the space of reconstructed quantities

[1]Two initial and six final state particles are described by 32 variables, but four energy-momentum conservation laws and four well-known masses (of the initial partons, $\ell$ and $\nu$) result in a trivial reduction to 24.

is described by ∼20 variables[2]. Moreover, the integrand has a complex structure because of the resonant nature of $t$ and $W$, and its efficient evaluation requires a highly nontrivial phase space sampling scheme. Inclusion of the processes beyond the leading order increases the phase space integral complexity even further. Since the number of integrals which has to be evaluated is usually quite large,[3] additional dimensionality reduction assumptions may be necessary for purely practical reasons: to simplify the reaction kinematics and to speed up the convergence of the integration procedures. The following ideas have been explored, in various combinations:

— Neglect some of the production channels (especially if their corresponding event fractions $f_i$ are expected to be small).

— Assume that some or all of the partons are on shell.

— Use tree-level, leading order matrix element and phase space.

— Further simplify the matrix element. Such simplifications might use narrow width approximations for the resonances, ignore spin correlations in pair production, consider only valence quarks in the production mechanisms, *etc.*

— Assume that initial partons have zero transverse momentum (hadron beams implied).

— Assume that some event variables can be perfectly measured in the detector (*i.e.,* transfer functions for these variables are represented by Dirac delta functions).

Each of these assumptions reduces the quality of the process statistical model and results in either degraded precision of a parameter estimator (in case the purpose of the analysis is a parameter measurement) or diminished power of a signal discriminant. Naturally, assumptions of this kind should be avoided unless either their effect could be shown to be negligible or the practical problems could not be overcome by other means. A number of techniques could be employed to increase the calculation speed without sacrificing the fidelity of the statistical model. These techniques include optimization of phase space and channel sampling [3, 4], use of multidimensional integration algorithms with better convergence properties than the standard Monte Carlo [5, 6, 7, 8], and speeding up the evaluation of the integrand (*e.g.*, by tabulating detector transfer functions for fast lookup, or by factorizing the integrand into terms which depend on different parameters so that some terms do not have to be recalculated for every value of $\mathbf{a}$).

## 3 Alternative Formulations of the Method

The first [1, 2] and some subsequent (*e.g.,* [9]) applications of the ME approach in HEP data analysis utilized a somewhat different formulation of the method. Expressed with the notation introduced in Section 1, the "extended likelihood" for the parameter $\mathbf{a}$ was postulated to be

$$L_{\,\mathrm{ext}}(\mathbf{a}) = e^{-N\int_Y \overline{P}_{\,\mathrm{ev}}(\mathbf{y}|\mathbf{a})\,d\mathbf{y}} \prod_{k=1}^{N} \overline{P}_{\,\mathrm{ev}}(\mathbf{y}_k|\mathbf{a}), \tag{3}$$

[2]Three momentum components for $\ell$ and each jet, and transverse distance of the secondary vertex to the beam axis for each jet. $\nu$ escapes undetected. Two additional variables should be included in the count if the missing transverse energy is used for event selection.

[3]For example, a recent measurement of the top quark mass [17] reports using a sample of 1,087 $t\bar{t}$ candidate events, while the likelihood is calculated on a $32 \times 26$ rectangular grid of parameter values. The number of Monte Carlo events which has to be processed in order to calibrate such a measurement is typically couple orders of magnitude higher.

where $N$ is the number of events in the data sample, and $\overline{P}_{\rm ev}(\mathbf{y}|\mathbf{a})$ stands for the probability to observe $\mathbf{y}$ without the requirement that $\int_Y \overline{P}_{\rm ev}(\mathbf{y}|\mathbf{a})\,d\mathbf{y}$ is normalized to 1. Although this may not be immediately obvious, this formula actually presumes that $\overline{P}_{\rm ev}(\mathbf{y}|\mathbf{a})$ contains an additional parameter: a freely floating normalization factor. Likelihood maximization with respect to that factor restores proper normalization of the $\overline{P}_{\rm ev}(\mathbf{y}|\mathbf{a})$, as illustrated in [1].

The utility of this particular technique is not clear: the normalization integral $\int_Y \overline{P}_{\rm ev}(\mathbf{y}|\mathbf{a})\,d\mathbf{y}$ is needed anyway in order to calculate the likelihood, and the parameter estimation requires an extra minimization step (typically, $-\ln L(\mathbf{a})$ is minimized in order to maximize the likelihood). There is, however, a more useful extended likelihood formulation which can actually exploit the correlation, if any, between the observed number of events and the probability density shape. A detailed description of this formulation can be found in [10].

A more substantial difference between Eq. 2 and the channel probability expressions used, *e.g.,* in [11, 12, 13, 14] consists in the explicit inclusion of the phase space efficiency term $\epsilon_i(\mathbf{x},\mathbf{a})$ in Eq. 2. It is indeed possible to write

$$P_i(\mathbf{y}|\mathbf{a}) = \frac{\Omega(\mathbf{y})}{\sigma_i(\mathbf{a})A_i(\mathbf{a})} \int_{X_i} W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})\,|M_i(\mathbf{x},\mathbf{a})|^2\,T_i(\mathbf{x},\mathbf{a})\,d\mathbf{x} \tag{4}$$

without this efficiency term, assuming transfer functions $W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})$ of a different kind. In order to maintain proper likelihood normalization, these transfer functions must satisfy a different normalization condition: $\int_Y W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})\,d\mathbf{y} = 1$ for every $i$, $\mathbf{x}$, and $\mathbf{a}$. It is obvious that the integrals in Eqs. 2 and 4 will result in the same likelihoods calculated for an arbitrary data sample in case $W_i'(\mathbf{y}|\mathbf{x},\mathbf{a}) = W_i(\mathbf{y}|\mathbf{x},\mathbf{a})\,\epsilon_i(\mathbf{x},\mathbf{a})$. This relationship can be multiplied by $\Omega(\mathbf{y})$ and integrated over $\mathbf{y}$. Together with the respective normalization conditions for $W_i$ and $W_i'$, this leads to an intuitive definition of phase space efficiency in terms of $W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})$ and $\Omega(\mathbf{y})$:

$$\epsilon_i(\mathbf{x},\mathbf{a}) = \int_Y \Omega(\mathbf{y}) W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})\,d\mathbf{y}. \tag{5}$$

There is, however, a crucial difference between functions $W_i(\mathbf{y}|\mathbf{x},\mathbf{a})$ and $W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})$: in order to completely define $W_i(\mathbf{y}|\mathbf{x},\mathbf{a})$, one has to know the function values only for those $\mathbf{y}$ for which $\Omega(\mathbf{y}) = 1$. On the other hand, the functions $W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})$ must be defined for *every* $\mathbf{y}$ so that their normalization integrals can be evaluated. As it will be discussed in the next section, no reliable algorithm has been identified so far for constructing such transfer functions — they suffer from *underspecification* in the $\Omega(\mathbf{y}) = 0$ region.

For the remainder of this article, the following terminology will be employed. Transfer functions which satisfy the normalization condition $\int_Y W_i'(\mathbf{y}|\mathbf{x},\mathbf{a})\,d\mathbf{y} = 1$ will be called "type I". Historically, these transfer functions were introduced first. Functions normalized by $\int_Y \Omega(\mathbf{y}) W_i(\mathbf{y}|\mathbf{x},\mathbf{a})\,d\mathbf{y} = 1$ will be designated "type II". Finally, the "type III" transfer functions will be formally defined via the type II functions by the expression $W_i''(\mathbf{y}|\mathbf{x},\mathbf{a}) \equiv W_i(\mathbf{y}|\mathbf{x},\mathbf{a})\,\epsilon_i(\mathbf{x},\mathbf{a})$. These functions are normalized by

$$\int_Y \Omega(\mathbf{y}) W_i''(\mathbf{y}|\mathbf{x},\mathbf{a})\,d\mathbf{y} = \epsilon_i(\mathbf{x},\mathbf{a}). \tag{6}$$

Although this equation looks almost identical to Eq. 5, its logic is inverted: instead of using the transfer function to determine the phase space efficiency, the efficiency is used to impose the transfer function normalization. Type III transfer functions defined in this manner do not suffer from the type I underspecification problem, and these are the correct functions to use instead of $W_i'$ inside the channel probability integral represented by Eq. 4.

## 4 Transfer Function Modeling

In order to obtain type II (or any other) detector transfer functions directly from data, one has to solve the inverse problem for $W_i(\mathbf{y}|\mathbf{x},\mathbf{a})$. Unfortunately, this is feasible only when the phase space $X_i$ is sufficiently restricted and the background contamination is minimal. Such conditions can indeed be found in test beams and in certain low-background processes on $e^+e^-$ colliders, but $pp$ or $p\bar{p}$ collider environments are not conductive to this approach. Therefore, transfer function derivation procedures utilized so far in HEP ME analyses always relied upon a reasonably well-tuned detector simulation software package. For a given channel, such a package starts with $\mathbf{x}$ and provides either a simulated $\mathbf{y}$ or an indication that the event does not fall inside the $\mathbf{y}$ region for which $\Omega(\mathbf{y}) = 1$.

Computationally, it is not feasible to estimate the transfer function by simulation in the process of evaluating Eq. 2 integral. For modern HEP experiments, it takes a few tens of CPU seconds to simulate particle detector response to an event[4]. Since at least a few hundred events are needed to form an estimate of the probability density for each phase space point and the number of phase space points per integral can be about $10^5$, the required CPU load is prohibitive. Instead, one constructs the transfer functions from pools of pre-simulated Monte Carlo (MC) events by introducing reasonable assumptions about the transfer function properties. The main assumptions are:

— To the first order, the full event transfer function can be factorized into the product of transfer functions for individual physics objects. Small corrections, *e.g.,* for nearby jets, can be introduced on top of this factorization.

— The transfer functions are piecewise continuous together with their first few derivatives. The discontinuity points are known in advance (for example, at the junctions of separate detector subsystems).

Usually, the detector response to jets is the most difficult transfer function factor to model. A significant fraction of ME analyses performed so far attempted to construct type I jet transfer functions by following the "original recipe" outlined in [1]. According to this recipe, jets are generated by MC with the same process as the process under study (this makes sense for factorized transfer functions without additional corrections, as energy response for most jet reconstruction algorithms does exhibit a mild dependence on jet multiplicity). The selection of events is performed either according to $\Omega(\mathbf{y})$ used for the data or according to a different $\Omega'(\mathbf{y})$ which corresponds to modified, "relaxed"[5] selection criteria. Simulated jets are matched to Monte Carlo partons using an angular matching criterion in the $\eta$-$\varphi$ space. The jet angular resolution is assumed to be perfect (modeled by Dirac delta functions), while a normalized probability density function (a double Gaussian is common) is fitted to a distribution of some energy-like *jet* quantity (transverse momentum or energy) using the same (or sometimes different) energy-like *parton* quantity as the explanatory variable (predictor). Some functional form (usually linear) is assumed for the parameters of the fitted density expressed in terms of the predictor. This whole procedure thus fits a nonlinear regression model in which the shape of the response distribution is constrained by its assumed functional form and depends on the predictor.

As it turns out, this recipe has a serious deficiency and it can't possibly produce fully correct type I transfer functions. The problem is that the Monte Carlo events in the fits are preselected,

[4] Typical for ATLAS and CMS detector simulation packages.

[5] $\Omega'(\mathbf{y}) = 1$ for every $\mathbf{y}$ for which $\Omega(\mathbf{y}) = 1$. In addition, $\Omega'(\mathbf{y}) = 1$ for some $\mathbf{y}$ for which $\Omega(\mathbf{y}) = 0$.

and only events with those $\mathbf{y}$ for which $\Omega(\mathbf{y}) = 1$ (or $\Omega'(\mathbf{y}) = 1$, depending on the details) are used. Thus the fitted MC events do not populate the complete $Y$ space on which type I transfer function normalization is defined. Such event samples can instead be used to fit type II transfer functions, but the recipe calls for a wrong functional form which can't represent a sharp cutoff in $\mathbf{y}$. Even if the sample selection criteria are relaxed as much as possible, the inherent inefficiencies of jet reconstruction algorithms would introduce an effective selection $\Omega''(\mathbf{y})$ which should not be ignored.

Defective transfer functions derived according to this recipe affect probabilities of events which contain low-energy jets near the $\mathbf{y}$ cutoff but do not invalidate the overall measurement results. The saving grace is provided by the analysis final calibration procedure which corrects for biases and pulls. However, just as unnecessary dimensionality reduction assumptions, such functions degrade either the measurement precision or the power of the signal discriminant.

Other methods of transfer function derivation have been developed. In Ref. [15] type I transfer functions were constructed by an altered procedure. The deficiency of the original recipe was noted and the $\Omega'(\mathbf{y}) = 1$ region was used to fit just the shapes of the transfer function densities but not to fix their normalization. The transfer functions were then extrapolated into the $\Omega'(\mathbf{y}) = 0$ region devoid of events. This method results in more appropriate functional shapes and makes the transfer function normalizable in the whole $Y$ space, but at the cost of introducing another problem: that of extrapolation ambiguity. The fatter the extrapolated tail inside $\Omega'(\mathbf{y}) = 0$, the smaller the function values become inside $\Omega'(\mathbf{y}) = 1$ due to the imposed normalization condition. For this method, the effect of transfer function mismodeling is difficult to assess. Nevertheless, this is undoubtedly an improvement upon the original recipe.

The authors of Ref. [16] explicitly build type II transfer functions using appropriate functional shapes which can model the $\Omega(\mathbf{y})$ cutoff. However, these functions are then used as if they were type I, inside Eq. 4. The authors demonstrate that in this case the overall probability normalization integral (called "observable cross section" in their paper) becomes independent of detector acceptance. Instead of raising the red flag, this property appeals to the authors because detector-related parameters no longer enter the likelihood normalization. It should be obvious to the reader of this note that this "simplification" of the likelihood comes at the price of an inappropriate transfer function model which distorts the statistical description of the process and increases the measurement error.

A proper construction of type III jet transfer functions was carried out in Ref. [17] (at the time of this writing, this work stands as the most precise single measurement of the top quark mass). The authors model both transverse momentum and angular jet response with nonparametric statistical techniques. They also employ a nonparametric estimate of the phase space efficiency to normalize their transfer functions in the $\Omega(\mathbf{y}) = 1$ region. The resulting transfer functions are then used to calculate event observation probabilities according to Eq. 4.

## 5 Practical Consequences

It should be apparent from the preceding discussion that an estimate of the phase space efficiency $\epsilon_i(\mathbf{x}, \mathbf{a})$ must be constructed in order to reliably calculate the probability $P_i(\mathbf{y}|\mathbf{a})$ to find an event with observables $\mathbf{y}$. The calculation can utilize either type II transfer function in combination with Eq. 2 or type III transfer function in combination with Eq. 4. Although these approaches are technically equivalent, the one which utilizes Eq. 2 is conceptually easier to understand because in

that equation $\epsilon_i(\mathbf{x}, \mathbf{a})$ has a very straightforward probabilistic interpretation.

The $\epsilon_i(\mathbf{x}, \mathbf{a})$ estimate can be obtained from the usual Monte Carlo event samples generated for analysis calibration purposes.[6] It is important that these event samples *are not filtered*, as often done in order to conserve disk space. This is because reconstruction of the $\mathbf{x}$-dependent efficiency denominator ($\epsilon$ = events accepted/events total) becomes difficult or impossible when a fraction of events in the sample is discarded based on the value of $\mathbf{y}$.

Finally, it should be pointed out that both type II and type III transfer functions can be constructed using "relaxed" MC event selection criteria (*e.g.*, in order to simplify normalization of the transfer functions when certain types of instrumental parameters, such as the overall jet energy scale, are scanned). In this case the corresponding "relaxed" phase space efficiencies should be derived and used inside Eqs. 2 or 6.

## References


[1] J.C. Estrada, "Maximal Use of Kinematic Information for the Extraction of the Mass of the Top Quark in Single-lepton $t\bar{t}$ Events at D0", Ph.D. thesis, University of Rochester (2001).

[2] D0 Collaboration, "A precision measurement of the mass of the top quark", Nature **429**, 638 (2004).

[3] P. Artoisenet, V. Lemaitre, F. Maltoni, and O. Mattelaer, "Automation of the matrix element reweighting method", arXiv:1007.3300v1 [hep-ph] (2010).

[4] R. Kleiss and R. Pittau, "Weight optimization in multichannel Monte Carlo", Comp. Phys. Commun. **83**, 141 (1994).

[5] H. Niederreiter, "Random Number Generation and Quasi-Monte Carlo Methods", Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992).

[6] R. Kleiss, A. Lazopoulos, "Error in Monte Carlo, quasi-error in Quasi-Monte Carlo", Comp. Phys. Commun. **175**, 93 (2006).

[7] J. S. Severino, E. J. Allen, H. D. Victory, "Acceleration of quasi-Monte Carlo approximations with applications in mathematical finance", Appl. Math. and Comput. **148**, 173 (2004).

[8] J. Dick, "Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands", arXiv:1007.0842v3 [math.NA] (2010).

[9] F. Canelli, "Helicity of the W Boson in Single-lepton $t\bar{t}$ events", Ph.D. thesis, University of Rochester (2003).

D0 Collaboration, "Helicity of the W boson in lepton + jets $t\bar{t}$ events", Phys. Lett. B **617**, 1 (2005).

[10] R. Barlow, "Extended maximum likelihood", Nucl. Instr. Meth. Phys. Res. A **297**, 496 (1990).


[6] A number of statistical techniques can be employed for this purpose. Local quadratic logistic regression is the personal favorite of the author.

[11] D0 Collaboration, "Measurement of the top quark mass in the lepton + jets final state with the matrix element method", Phys. Rev. D **74**, 092005 (2006).

[12] CDF Collaboration, "Precise Measurement of the Top-Quark Mass in the Lepton+Jets Topology at CDF II", Phys. Rev. Lett. **99**, 182002 (2007).

[13] D0 Collaboration, "Observation of Single Top-Quark Production", Phys. Rev. Lett. **103**, 092001 (2009).

[14] CDF Collaboration, "Observation of Electroweak Single Top-Quark Production", Phys. Rev. Lett. **103**, 092002 (2009).

[15] J. Freeman, "A Measurement of the Top Quark Mass in 1.96 TeV Proton-Antiproton Collisions Using a Novel Matrix Element Method", Ph.D. thesis, University of California, Berkeley (2007).

CDF Collaboration, "Top quark mass measurement in the lepton plus jets channel using a modified matrix element method", Phys. Rev. D **79**, 072001 (2009).

[16] F. Fiedler, A. Grohsjean, P. Haefner, and P. Schieferdecker, "The Matrix Element Method and its Application to Measurements of the Top Quark Mass", arXiv:1003.1316v2 [hep-ex] (2010).

[17] P.J. Lujan, "Precision Measurement of the Top Quark Mass in the Lepton + Jets Channel Using a Matrix Element Method with Quasi-Monte Carlo Integration", Ph.D. thesis, University of California, Berkeley (2009).

CDF Collaboration, "Top Quark Mass Measurement in the lepton+jets Channel Using a Matrix Element Method and in situ Jet Energy Calibration", Phys. Rev. Lett. **105**, 252001 (2010).